Sonoma Technology, Inc.

1360 Redwood Way, Suite C
Petaluma, CA 94954
707/665-9900
FAX 707/665-9800
www.sonomatech.com

DEVELOPMENT OF GRIDDED SPATIAL ALLOCATION FACTORS FOR THE STATE OF TEXAS

FINAL REPORT 3
STI-900570-2114-FR3

By:
Tami H. Funk
Patricia S. Stiefer
Lyle R. Chinkin
Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169

Prepared for:
Texas Natural Resource Conservation Commission
12100 Park 35 Circle
Austin, TX 78753

August 31, 2001

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This technical memorandum discusses sources of data and describes in detail methodologies employed to develop gridded spatial allocation factors for the state of Texas. These spatial allocation factors, which will be used to geographically distribute area and non-road mobile source emissions, were developed from *spatial surrogate data*. Spatial surrogates are economic, demographic, and land cover patterns that vary geographically.

Gridded spatial allocation factors for a 2000 base-year were developed for the entire state of Texas as well as portions of Louisiana, Arkansas, and Oklahoma. The grid domain for which spatial allocation factors were developed is defined in detail in the *Protocol for Ozone Modeling of the Houston/Galveston Area for the 2002 Early Assessment (Phase I of the 2004 Mid-Course Review)* and covers all of eastern Texas and portions of Louisiana, Arkansas, and Oklahoma. The domain is resolved to a grid cell size of 2-km x 2-km with nested 1-km x 1-km grids covering the Houston-Galveston, Beaumont-Port Arthur, Dallas-Fort Worth, Austin, San Antonio, Longview-Tyler-Marshall, Corpus Christi, and Victoria non-attainment/near-non-attainment areas. The 2-km grid and 1-km sub-grid domains are shown in **Figure 1-1**. **Table 1-1** contains the definition of each grid.

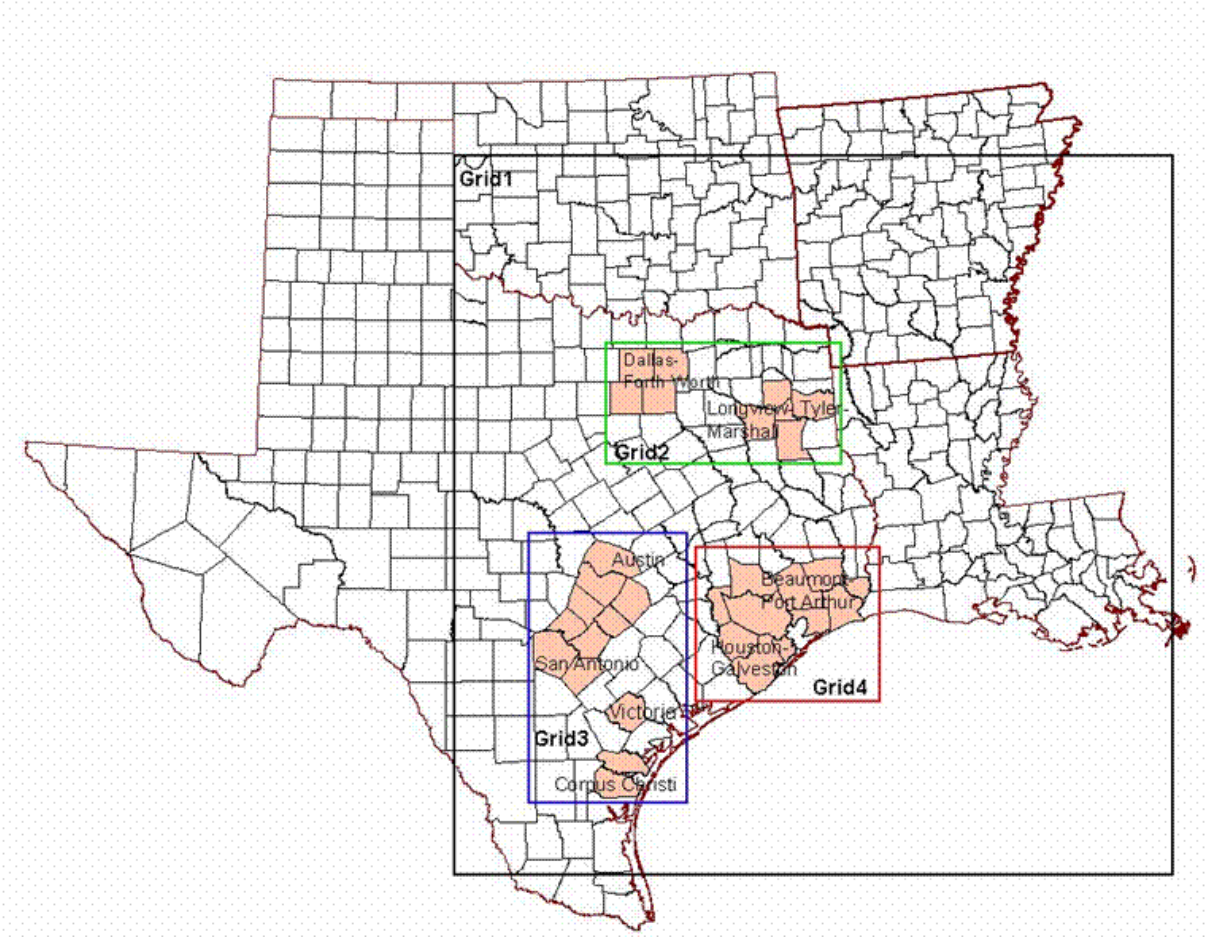


Figure 1-1. Depiction of the 2-km x 2-km grid domain and 1-km x 1-km sub-grids for which spatial allocation factors were developed.

Table 1-1. Projection and definition of grids for which spatial allocation factors were developed.

Projection:	Lambert Conformal Conic			
Parameters:	1st Standard Parallel: 30.0 degrees 2nd Standard Parallel: 60.0 degrees Longitude of Central Meridian: -100.0 degrees Latitude of Projection Origin: 40.0 degrees False Easting: 0.0 False Northing: 0.0			
Units:	meters			
Spheroid:	Radius: 6,370,997 meters			
	Grid 1	**Grid 2**	**Grid 3**	**Grid 4**
Origin:	0, -1476000	220000, -880000	108000, -1372000	350000, -1226000
Resolution:	2 km	1 km	1 km	1 km
Extent:	522 x 522	176 x 342	232 x 392	268 x 226

A thorough review was performed to identify sources of regional, local, and statewide surrogate data appropriate for use in the development of spatial allocation factors for the Texas surrogate project; and a Geographic Information System (GIS)-based approach was used to create a processing scheme to grid and display the surrogate and spatial allocation factor data.

The remainder of this technical memorandum is divided into two sections. Section 2 discusses the sources of underlying surrogate data used in this project, and Section 3 provides a discussion of the methodologies and processing steps used to develop the spatial allocation factors.

2. SOURCES OF SURROGATE DATA

Land use and land cover data are used to develop spatial allocation factors that are associated with specific land uses (e.g., agriculture, residential areas, and recreation areas). For emissions sources associated with residential locations and industrial, commercial, and agricultural areas, the new 30-meter resolution United States Geologic Survey (USGS) land cover and land use data were used to develop spatial allocation factors. For emissions sources associated with population density, data from the U.S. Census 2000 were used. For emissions from sources such as gasoline service stations and drycleaners, facility locations were used to develop spatial allocation factors. A thorough review of available data was carried out to identify existing sources of surrogate data to update and improve surrogates used in the past. Consideration was given to the resolution, vintage, and representativeness of surrogate data compared to emissions sources. Several sources of surrogate data were identified including

- Local transportation planning agencies (TPAs)
- Air pollution control agencies
- State and federal agencies
- U.S. Census Bureau
- Commercially available data sources

A technical memorandum, dated June 15, 2001, was delivered to the TNRCC containing a detailed discussion of the sources of identified surrogate data as well as recommendations for surrogate development.

Several sources of land use and land cover data were identified and evaluated for use in surrogate development. **Table 2-1** contains a summary of the land use and land cover data sets that were used to develop spatial allocation factors.

Table 2-1. Sources of land use and land cover data used to develop spatial allocation factors.

Data Source	Parameter	Resolution	Vintage	Coverage
United States Electronic Yellow Pages (*ProCD Select Phone*)	Autobody shops, dry cleaners, restaurants, gas stations, and marine coatings facilities	Address locations	Late 1997	TX
Environmental Systems Research Institute	Airports, military bases	Polygon coverages	1997	TX
U.S. Census Bureau (ESRI ADOL version)	Water bodies	Polygon coverages	2000	TX
U.S. Census Bureau	Urban boundaries	Polygon coverages	1990	TX, LA, AR, OK
United States Geological Survey	New land use and land cover data	Image data	1997- 2000	TX, LA, AR, OK

Table 2-1. Sources of land use and land cover data used to develop spatial allocation factors.

Data Source	Parameter	Resolution	Vintage	Coverage
Bureau of Transportation Statistics	Ports, shipping lanes, canals, railroads	Coordinate locations and line coverages	Publication date is 2000; source date varies	TX, LA
U.S. Census Bureau	Population	Block level polygons	2000	TX, LA, AR, OK
Texas Railroad Commission	Oil and gas well locations[a]	Coordinate locations	1998	TX

[a] Data for oil and gas well locations was obtained by TNRCC, however spatial allocation factors were not developed as part of this work due to proprietary data issues.

2.1 ASSIGNMENT OF SURROGATES TO EMISSIONS SOURCE CATEGORIES

Spatial allocation factors are used to spatially disaggregate countywide emissions into grid cells. Each identified surrogate must be assigned to a source category (or multiple categories) in the emission inventory. This assignment provides a cross reference between the spatial allocation factors and the emission inventory categories. A total of 24 surrogates were developed as part of this project.

3. DEVELOPMENT OF GRIDDED SPATIAL ALLOCATION FACTORS

A GIS-based approach was used to develop gridded surrogate data and spatial allocation factors. The GIS-based approach consists of the following steps:

1. Manipulation of all geographic data sets into the required projection and format(s) for the GIS and processing
2. Manipulation of all tabular data sets into a format required for creating gridded surrogates and spatial allocation factors
3. Spatial disaggregation of surrogate coverages into grid cells using ArcInfo and customized processing scripts written in Arc Macro Language (AML)
4. Use of Microsoft Access databases to calculate the gridded surrogates and spatial allocation factors for each set of surrogate data
5. Use of ArcView to display gridded surrogate and spatial allocation factor data contained in Access databases
6. Quality assurance of the gridded surrogate and spatial allocation factor files
7. Preparation of final spatial allocation factor files for TNRCC

The remainder of Section 3 provides a detailed discussion of each of the seven processing steps employed to develop the gridded surrogates and spatial allocation factor files.

3.1 OVERVIEW OF THE GIS-BASED DATA PROCESSING SCHEME

The GIS-based data processing scheme is comprised of the following five tiers:

Tier 1—Preparation of Surrogate Basemaps: The raw geographic surrogate data files were processed (i.e., re-projected to the grid domain and cleaned) to generate ArcInfo coverage basemaps.

Tier 2—Level_1_Databases: Level_1_Databases contain the raw tabular population data from the U.S. Census 2000 for Texas, Louisiana, Oklahoma, and Arkansas. The function of the Level_1_Database is to place the raw Census data into a standardized format for processing in the Level_2_Databases.

Tier 3— Level_2_Databases: Level_2_Databases contain all gridded geographic and tabular data necessary for calculating the gridded surrogate and gridded spatial allocation factor files.

Tier 4— Data Displays: ArcView project files were created to display and quality assure all of the gridded surrogate and spatial allocation factor data created by the Level_2_Databases.

Tier 5—Master Databases: The Master Databases contain all of the tabular spatial allocation factor files for each grid in a standard format.

The first level of processing involves manipulation and formatting of all geographic surrogate data. The next level of data processing involves preparing the Level_1_Database containing the raw U.S. Census 2000 data. The Level_1_Database is designed to pre-process the raw Census data files into a standardized format required by the Level_2_Databases. The Level_2_Databases perform the gridded surrogate and spatial allocation factor calculations and output files for all surrogate data sets. Displays of the gridded surrogate and spatial allocation factor data from the Level_2_Databases is contained in the GS_SAF_Displays directory. ArcView shapefiles are created for each gridded surrogate and spatial allocation factor data set. ArcView project files are created to display and quality assure data for each surrogate.

The final level of processing involves creating the Master Databases which contain all of the final spatial allocation factor files. The directories containing all TNRCC gridded surrogate project files are summarized in **Table 3-1**. The same general procedures were followed for processing data for each grid definition. Directories containing "_G2", "_G3", and "_G4" in their names correspond to data for grids 2, 3, and 4 (respectively).

When working with ArcInfo map files (.mxd) and ArcView project files (.apr), it is important to preserve the directory structure so that the map files can be opened. If files and/or directories are moved from their original locations, it may cause the map files to become unreadable. To preserve the directory structure on which the surrogate data were developed, the following directory path should be established at TNRCC: E:/Project_Files/900570_Grid_Surg. The directories and sub-directories should be consistent with those presented in Table 3-1.

Table 3-1. Directories containing all TNRCC gridded surrogate project files.

Directory Name	Contents
Aml_templates	Contains generic AML scripts used for this project.
Basemaps	Contains all surrogate basemaps, gridded basemaps, and AML gridding scripts for Grid_1 (see Figure 1-1). The Level 1 ArcInfo layers and map files are also contained in this directory.
Basemaps_G2	Contains all surrogate basemaps, gridded basemaps, and AML gridding scripts for Grid_2 (see Figure 1-1).
Basemaps_G3	Contains all surrogate basemaps, gridded basemaps, and AML gridding scripts for Grid_3 (see Figure 1-1).
Basemaps_G4	Contains all surrogate basemaps, gridded basemaps, and AML gridding scripts for Grid_4 (see Figure 1-1).
Documentation	Contains project documentation files, including this technical memorandum.

Table 3-1. Directories containing all TNRCC gridded surrogate project files.

Directory Name	Contents
Grid_1	Contains the ArcInfo coverage and ArcView Shapefile for Grid_1.
Grid_2	Contains the ArcInfo coverage and ArcView Shapefile for Grid_2.
Grid_3	Contains the ArcInfo coverage and ArcView Shapefile for Grid_3.
Grid_4	Contains the ArcInfo coverage and ArcView Shapefile for Grid_4.
Level_1_Databases	Contains a database with raw U.S. Census 2000 data (population only).
Level_2_Databases	Contains all the Level_2_Databases used to create gridded surrogate (GS) and spatial allocation factor (SAF) files for Grid_1. Also contains ArcInfo feature attribute tables (FAT_Files) and gridded data files used for Level 2 processing.
Level_2_Databases _G2	Contains all the Level_2_Databases used to create GS and SAF files for Grid_2. Also contains gridded data files used for Level 2 processing.
Level_2_Databases _G3	Contains all the Level_2_Databases used to create GS and SAF files for Grid_3. Also contains gridded data files used for Level 2 processing.
Level_2_Databases _G4	Contains all the Level_2_Databases used to create GS and SAF files for Grid_4. Also contains gridded data files used for Level 2 processing.
GS_SAF_Displays	Contains ArcView shapefiles and project files for each gridded surrogate and spatial allocation factor data set corresponding to Grid_1.
GS_SAF_Displays _G2	Contains ArcView shapefiles and project files for each gridded surrogate and spatial allocation factor data set corresponding to Grid_2.
GS_SAF_Displays _G3	Contains ArcView shapefiles and project files for each gridded surrogate and spatial allocation factor data set corresponding to Grid_3.
GS_SAF_Displays _G4	Contains ArcView shapefiles and project files for each gridded surrogate and spatial allocation factor data set corresponding to Grid_4.
Master Database	Final gridded spatial allocation factor files for Grid_1.
Master Database_G2	Final gridded spatial allocation factor files for Grid_2.
Master Database_G3	Final gridded spatial allocation factor files for Grid_3.
Master Database_G4	Final gridded spatial allocation factor files for Grid_4.

3.2 PREPARATION OF SURROGATE BASEMAPS

GIS data for the 2000 Census (at the block level), land use, and land cover categories were obtained from various sources in widely varying formats and map projections. The procedures for assembling this data into the surrogate basemaps, and some of the individual data sets processed, are described below. Each basemap was displayed in an ArcInfo map file (.mxd); and is contained in the following directory:
E:\Project_Files\900570_Grid_Surg\Basemaps\Level_1_Maps.

3.2.1 General Procedures

Following are the general steps for assembling the surrogate basemaps:

- Import the original GIS data into an ArcInfo coverage and re-project it to match the Grid_1 domain projection.
- Overlay the surrogate coverage with county boundaries to visually verify location.
- Identify and resolve any geometry or placement errors that are beyond acceptable limits.
- Add and populate attribute table fields for the county FIPS codes and standardize cross-reference codes for Census Block polygons.

3.2.2 Census Block Basemaps

STI obtained 2000 Census Block area boundaries for Texas, Arkansas, Louisiana, and Oklahoma. The GIS data were downloaded from the U.S. Census Bureau's Cartographic Boundary Files web site in ArcView Shapefile format. Because the Census Block data have a long block identifier, a simple numeric cross-reference field (XREF) was added to each statewide geographic data set. The XREF is a number from 1 to X which makes possible a standard numeric polygon identifier corresponding to each Census Block code.

3.2.3 Land Use and Land Cover Basemaps

GIS data for land use and land cover maps were obtained from the various sources listed in Table 2-1. This data varied widely in format and projection. Some original data sets required additional editing or filtering before the final basemap could be completed.

- Recreational use water. Water body polygons from the 2000 Census data for all counties in Texas were merged and manually edited in ArcView before being imported into ArcInfo. The edit process eliminated all dry lakes and retained all lakes and rivers with boat marinas.
- Facilities. Locations from the Electronic Yellow Pages (EYP) were geocoded to street addresses or zip code centroids. Sources collocated at a zip code centroid were randomly dispersed throughout the zip code boundary.

- Urban areas. The urban area polygons are based on 1990 U.S. Census Bureau's definitions for Urbanized Areas; the remaining land area outside a 1990 Urbanized Area is rural land. An Urbanized Area is only partly based on population density and consists of the total of the following areas:
  - One or more central places (either a central city within a Census-defined metropolitan area or a city having a population above a certain size)
  - Urban fringe (adjacent, contiguous territory having a density of at least 1,000 persons per square mile)

3.3 PROCESSING OF TABULAR DATA SETS

The Level_1_Databases contain the raw 2000 Census Block population data for all four states. The Level 1 Census database contains a standardized polygon XREF block identifier assigned to each record in the raw data. The raw Census Block codes are long alphanumeric strings. To simplify the Census Block identifiers, the XREF field was created to give each Block an integer value from 1 to X. Queries in the Level_1_Databases create the output tables, which are linked to the Level_2_Databases (Census_Blocks_XX files) along with the gridded GIS tabular files and geographic feature attribute tables. The gridded surrogates and spatial allocation factors are calculated in the Level_2_Databases.

3.4 GRIDDING OF SURROGATE BASEMAPS

Development of the gridded surrogate and spatial allocation factor files was carried out using ArcInfo scripts written in Arc Macro Language (AML) and the Level_2_Databases. Each surrogate coverage file was spatially disaggregated into grid cells using AML scripts to carry out a sequence of ArcInfo processing steps. Each of the surrogate coverage files was disaggregated to grid cells using the following three types of spatial overlay functions:

- Point-in-polygon – overlaying point location data and the grid domain
- Line-in-polygon – overlaying line coverage and the grid domain
- Polygon-in-polygon – overlaying polygon coverage and the grid domain

Figure 3-1 provides a conceptual illustration of the gridding process. All AML gridding scripts contain the name of the AML file followed by "_gso". These scripts are located in each of the surrogate basemap directories. The AML scripts are consistent for each surrogate data set. Each AML script prompts ArcInfo to carry out the following processes:

- An overlay is performed which combines the geographic features of the surrogate coverage with the grid coverage and creates a new coverage designated by "_g" in the filename indicating that the surrogate coverage has been gridded.
- The gridded surrogate coverage file (_g) is opened and all records with no spatial overlap between the surrogate and the grid are removed from the file (this feature was added to reduce extraneous data from each file). A new coverage file is then generated containing a "_gs" in the filename indicating that the surrogate coverage has been gridded and only the pertinent records have been selected.

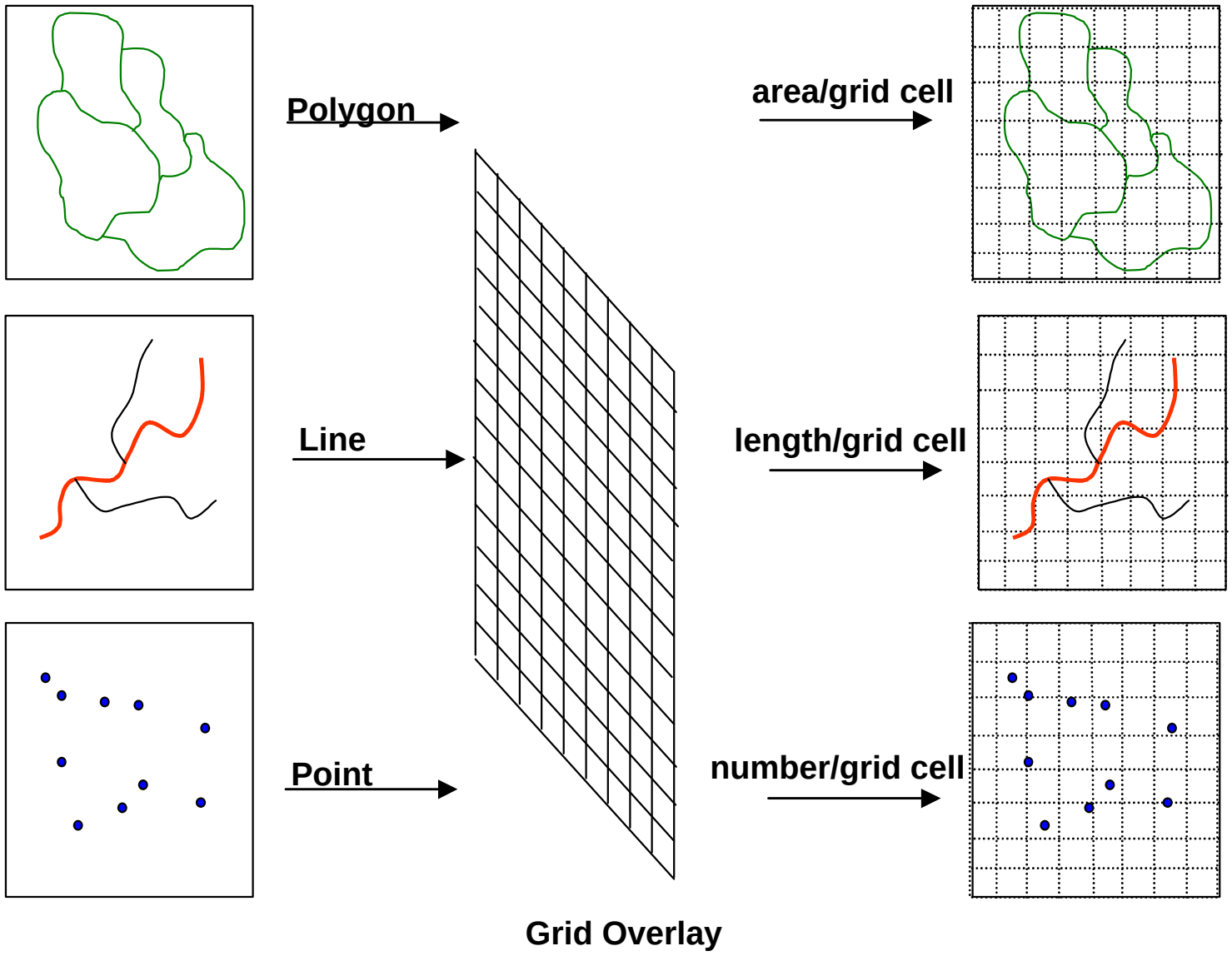


Figure 3-1. Illustration of disaggregating geographic objects into grid cells.

- The gridded and selected coverage file (_gs) is opened and a new item (or field) is added to the coverage attribute table. The new field is populated with a unique identification code consisting of the grid cell identifier and the surrogate feature identifier.
- The density of surrogate features (e.g., area, length, or number) contained in each grid cell is calculated and an output file is generated with "_gso" in the filename indicating that the file has been gridded, selected, and output.

Figure 3-2 includes an example AML gridding script with annotated text next to each command line of the script. The gridded output files were then imported into the Level_2_Databases where they were combined with tabular data (when applicable), and the gridded surrogate and spatial allocation factor files were calculated.

Because the state of Texas is so large and contains hundreds of thousands of high resolution Census Blocks, the statewide Census Block coverage was split into five pieces (TX1-TX5). **Figure 3-3** shows the extent of each TX coverage. Splitting the statewide coverage into five pieces resulted in much quicker processing times and data files of reasonable size. Gridded surrogate data for the five regional coverages were merged into one uniform data file following the Level 2 processing.

3.5 CALCULATION OF GRIDDED SURROGATES AND SPATIAL ALLOCATION FACTORS

The Level_2_Databases were used to combine the gridded geographic surrogate data with tabular surrogate data (when applicable) and to calculate the gridded surrogate and spatial allocation factor data files. The Level_2_Databases were named based on the type of data they contain. There are four general categories of Level_2_Databases including

- Databases containing Census Block data
- Databases containing land use and land cover data
- Databases containing transportation routes (i.e., roads and railroads)
- Databases containing area source facility location data

3.5.1 Level 2 Databases

The following are Level 2 databases:

- Airports.mdb (Airport surrogate data)
- Census_Blocks_AR.mdb (Census Block data for Arkansas)
- Census_Blocks_LA.mdb (Census Block data for Louisiana)
- Census_Blocks_OK.mdb (Census Block data for Oklahoma)
- Census_Blocks_TX1.mdb (Census Block data for Texas_1)
- Census_Blocks_TX2.mdb (Census Block data for Texas_2)
- Census_Blocks_TX3.mdb (Census Block data for Texas_3)
- Census_Blocks_TX4.mdb (Census Block data for Texas_4)
- Census_Blocks_TX5.mdb (Census Block data for Texas_5)
- Census_Blocks_TX_Merge.mdb (Census Block data merged for Texas)

- Counties.mdb (County area surrogate data)
- EYP.mdb (Facility location surrogate data)
- MilitaryBases.mdb (Military base surrogate data)
- Railroad.mdb (Railroad surrogate data)
- Rural.mdb (Rural area surrogate data)
- Shipping.mdb (Shipping lane surrogate data)
- Urban.mdb (Urban area surrogate data)
- WaterBodies.mdb (Water body surrogate data)
- USGS_LandUse.mdb (USGS land use surrogate data)

Level 2 Census Block databases were constructed with links to the tables in the corresponding Level 1 Census database. The population surrogate data contained in the Level 1 Census Block database was used in the Level_2_Databases to create the gridded surrogate and spatial allocation factor files for population. The Level_2_Databases combine the output of the ArcInfo gridding processes with the tabular surrogate data in the Level_1_Databases. **Table 3-2** provides a summary of the tables and queries contained in the Level 2 Census Block databases.

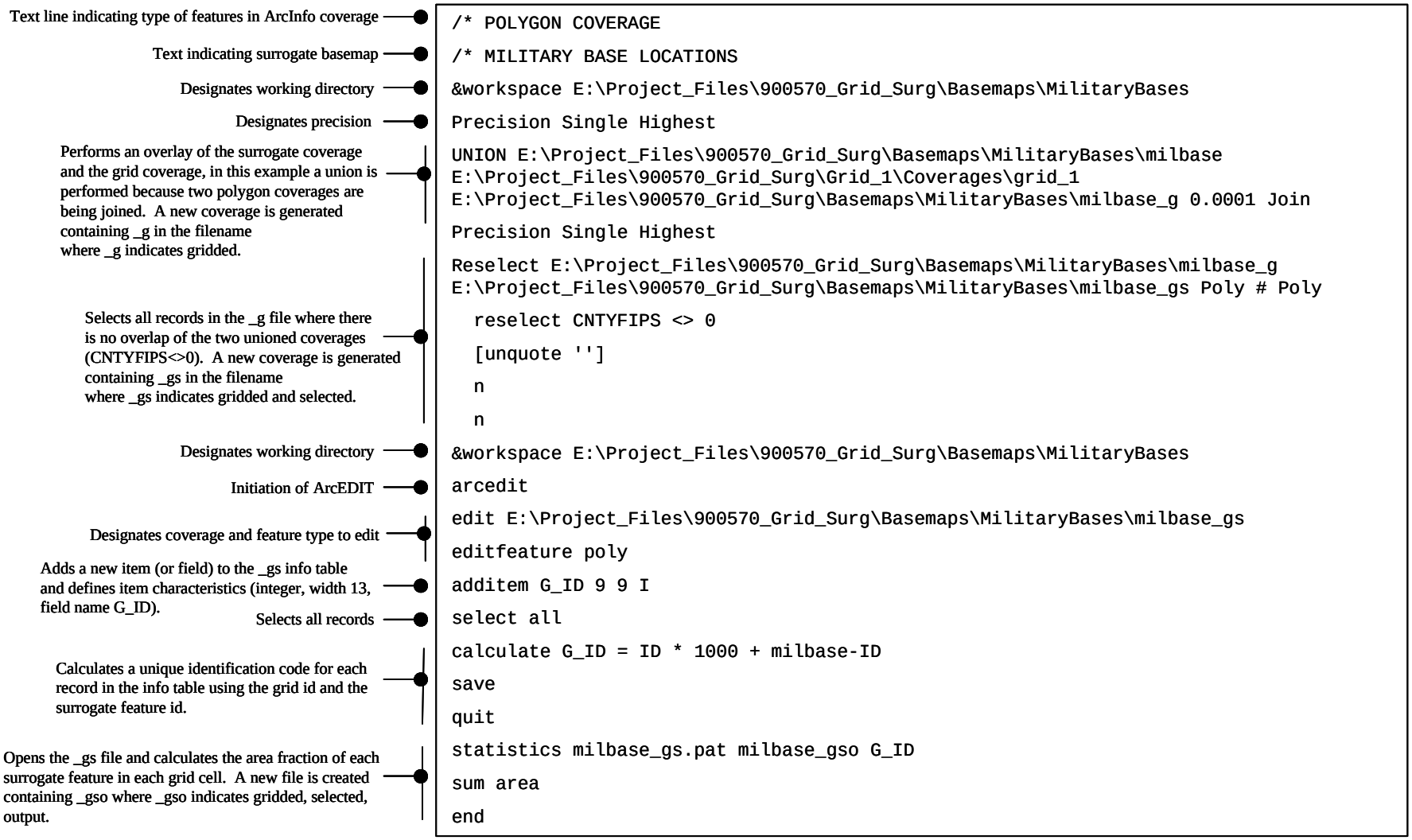


Figure 3-2. Example of an ArcInfo AML script used to carry out gridding processes.

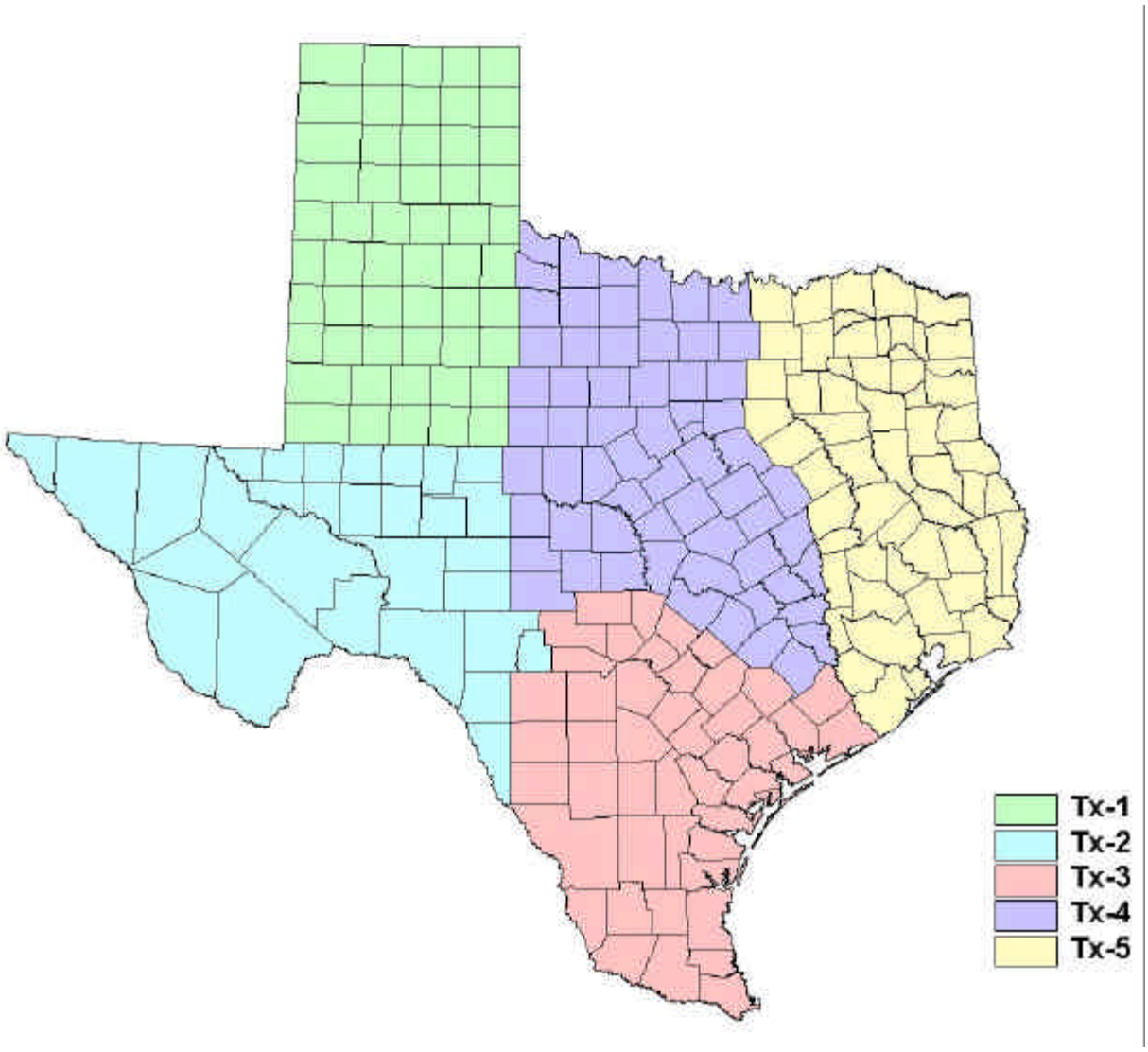


Figure 3-3. Five areas of Texas depicted for population data processing.

Table 3-2. Summary of tables and queries contained in the Level 2 Census Block databases.

Table Name	Description
GRIDDED_INPUT	Gridded surrogate data, output from ArcInfo AML gridding scripts. Contains the area fraction of each Census Block in each grid cell.
CNTY_TOTALS	Contains surrogate totals by county.
SSC_GS	Gridded surrogate file.
SSC_SAF	Gridded spatial allocation factor file.
Query Name	**Description**
Update_Gridded_Input (1)	Updates the gridded input file with GRID_ID and XREF, and calculates area in kilometers.
Update_Gridded_Input (2)	Updates the gridded input file with the county FIPS.
XREF_AREAS	Calculates the total area of each Census Block in each data set.
SSC_GS_CALC	A Make Table query that calculates the gridded surrogate file and creates SSC_GS table.
SSC_SAF_CALC	A Make Table query that calculates the gridded spatial allocation factor file and creates SSC_SAF table.

The first step in the Level_2_Database processing was to link the appropriate table from the Level 1 Census Block database. Once these tables were linked to the appropriate Level_2_Database, the gridded output file from the ArcInfo gridding process was imported into the database. The next step in the processing was to update the GRIDDED_INPUT file with three fields: (1) GRID_ID—the grid cell identifier; (2) XREF—the cross reference code representing each Census Block in the surrogate data set; and (3) AREA_KM2—calculates the area (in kilometers) of each Census Block in each grid cell. In order to do this, three new fields (with the appropriate field names) were added to the GRIDDED_INPUT file. Next, the Update_Gridded_Input (1) query was run which updates the new fields with the correct data values. Then, the Update_Gridded_Input (2) query was used to populate the GRIDDED_INPUT file with the county FIPS. To calculate gridded surrogates and spatial allocation factors, both the fraction of the Census Block residing in a grid cell and the total area of the Census Block must be known. The XREF_AREAS query calculates the total area of all Census Block zones.

The gridded surrogate files were calculated using the SSC_GS_CALC query. The SSC_GS_CALC query uses the GRIDDED_INPUT, XREF_AREAS, and Census Block population data table (linked from Level 1) tables to calculate the gridded surrogate values using the following general equation:

$$SSC\_GS = (AF_{(A)(g1)}/AT_{(A)})*SSC\_V_{(A)}$$

where:

SSC_GS = Gridded surrogate value for a given spatial surrogate code (SSC)
$AF_{(A)(g1)}$ = Area fraction of Census Block (A) in grid cell (1)

$AT_{(A)}$ = Total area of Census Block (A)
$SSC\_V_{(A)}$ = Spatial surrogate value for Census Block (A)

The gridded surrogate value is calculated by dividing the area fraction of a Census Block in a grid cell by the total area of the Census Block. The result is the fraction of the total Census Block in the grid cell. The surrogate data value corresponding to that Census Block is then multiplied by the area fraction. This computation results in the surrogate value per grid cell.

Spatial allocation factors are weighted values that indicate what fraction of the county total surrogate value resides in each grid cell. The gridded spatial allocation factors were calculated using the SSC_SAF_CALC query. The SSC_SAF_CALC query uses the GRIDDED_INPUT, XREF_AREAS, CNTY_TOTAL, and Census Block population data table (linked from Level 1) to calculate the gridded spatial allocation values using the following general equation:

$$SSC\_SAF = [(AF_{(A)(g1)}/AT_{(A)})*SSC\_V_{(A)}]/CT_{SSC}$$

where:

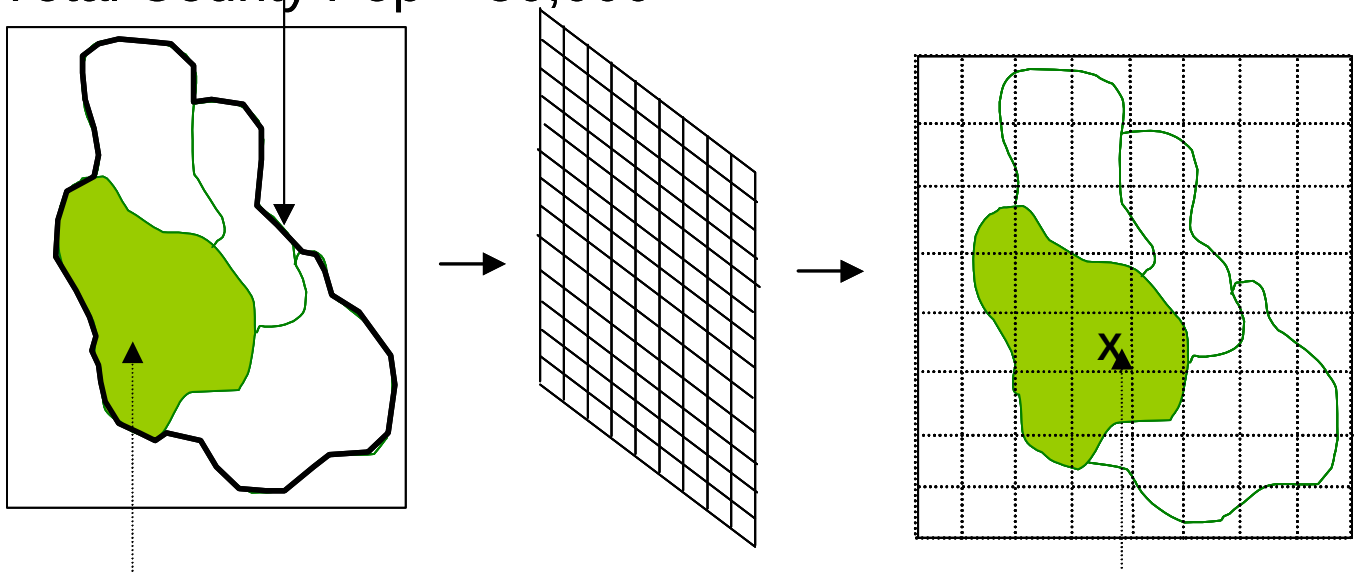
SSC_SAF = Gridded spatial allocation factor value for a given spatial surrogate code (SSC)
$AF_{(A)(g1)}$ = Area fraction of Census Block (A) in grid cell (1)
$AT_{(A)}$ = Total area of Census Block (A)
$SSC\_V_{(A)}$ = Spatial surrogate value for Census Block (A)
CT_{SSC} = County total spatial surrogate value

The gridded spatial allocation factors are calculated using the same equation for the gridded surrogates and then dividing the gridded surrogate value by the total surrogate value for the county. The result is the fraction of the county total surrogate value in the grid cell. **Figure 3-4** provides an illustrated example of how the spatial allocation factors are calculated.

The output of the gridded surrogate and spatial allocation factor queries are the SSC_GS and SSC_SAF tables containing gridded values for each of the demographic databases. These tables were then exported out of the Access databases as text files and displayed using ArcView (displays located in the GS_SAF_Displays directory).

Example Calculation: Population

Total County Pop = 50,000



Pop of Census Block = 5,000

Pop of Grid Cell (x) = 500

SAF = Population of Grid Cell / Population of County

= 500 / 50,000

= 0.01

1% of the county total population resides in grid cell (x)

Figure 3-4. Example of the spatial allocation factor calculation.

3.5.2 Level 2 Land Use and Land Cover Databases

The Level 2 land use and land cover databases use the output of the ArcInfo gridding processes to generate the gridded surrogate and spatial allocation factor files. The Level 2 land use and land cover databases are listed above (section 3.5.1). **Table 3-3** provides a generic summary of the tables and queries contained in the Level 2 land use and land cover databases.

Table 3-3. Summary of tables and queries contained in the Level 2 land use and land cover databases.

Table Name	Description
BasemapName_grid or GRIDDED_INPUT	Gridded surrogate data, output from ArcInfo AML gridding scripts. Contains the area, length, or point density of each geographic feature in each grid cell.
BasemapName_fat	ArcInfo coverage feature attribute table.
SSC_GS	Final gridded surrogate data file.
SSC_SAF	Final gridded spatial allocation factor data file.
Query Name	**Description**
Update_Gridded_Input (1)	Updates the gridded input file with GRID_ID and ArcInfo-id field, and calculates area or length (in kilometers), or count of features.
Update_Gridded_Input (2)	Updates the gridded input file with the county FIPS.
XREF_AREAS	Calculates the total area, length, or count of features in each data set.
SSC_GS_CALC	A Make Table query that calculates the gridded surrogate file and creates SSC_GS table.
SSC_SAF_CALC	A Make Table query that calculates the gridded spatial allocation factor file and creates SSC_SAF table.
CNTY_Totals	Calculates the surrogate totals by county.

The first step in the Level 2 land use and land cover database processing was to import the gridded data file from the ArcInfo gridding processes. The ArcInfo attribute table for each coverage was also imported into the database. The gridded input table was modified to include the following four fields: GRID_ID, ArcInfo-id, AREA_KM2 (or LENGTH_KM), and the county FIPS. Update queries 1 and 2 populate these fields with data values. The CNTY_Totals query calculates the surrogate totals by county. The SSC_GS_CALC query is a Make Table query that calculates gridded surrogate values by grid cell and creates the SSC_GS table. The SSC_SAF_CALC query is a Make Table query that calculates spatial allocation factors by grid cell and creates the SSC_SAF table. The spatial allocation factor file divides the gridded surrogate values by the county total surrogate value which results in a weighting factor representing the fraction of each surrogate in each grid cell based on the county total. All spatial allocation factors for each county add up to 1 (or 100 if expressed as a percent).

3.6 DISPLAY OF GRIDDED SURROGATES AND SPATIAL ALLOCATION FACTORS

The GS_SAF_Displays directory contains gridded surrogate and spatial allocation factor data displays in ArcView project file format. Each gridded surrogate and spatial allocation factor data file was exported from the Level_2_Databases as text files. Each gridded surrogate and spatial allocation factor data table was joined to a grid shapefile in ArcView to display the values of the individual grid cells across the domain. The result of the ArcView join was saved permanently as a shapefile. An ArcView project file (.apr) was created for each gridded surrogate and each spatial allocation factor display. The purpose of the data displays is to quality assure the gridded surrogate and spatial allocation factor files.

3.7 QUALITY ASSURANCE

Quality assurance protocols were carried out during each stage of data processing. Quality assessments of each surrogate basemap were made to confirm that the geographic location and content of each data set was correct within a specified range (as discussed in Section 3.2). The surrogate and spatial allocation factor data contained in the Level_2_Databases were reviewed, and intermediate queries were created, to verify that the sum of each county's total spatial allocation factor added to 1 (or 100 if expressed as a percent).

A general quality assurance procedure consisted of displaying the basemap, gridded surrogate data, and spatial allocation factor data for each gridded surrogate, or SSC. The basemap display was overlayed with the gridded display and the spatial allocation factor map to verify consistency among each level of data processing. When errors were found, corrections were made; and the quality assurance procedures were repeated until each stage of the data processing had been completed successfully. Once all spatial allocation factor files had been checked for quality and consistency, the Master Databases were assembled.

3.8 PREPARATION OF FINAL MASTER DATABASES

The following are Master Databases:

- GRIDDED_SAFS.mdb
- GRIDDED_SAFS_G2.mdb
- GRIDDED_SAFS_G3.mdb
- GRIDDED_SAFS_G4.mdb

The GRIDDED_SAFS databases contain the gridded spatial allocation factor data for all spatial surrogates and counties for each state. The SSC_SAF tables in the Level_2_Databases are linked to the Master Databases. The Master Databases contain queries to merge all of the final spatial allocation factor data files into one table. There is one Master Database containing data for each grid definition.